

## Group name: Solitude Ensemble

### Group member's details

---

Name: Parthkumar Sutariya  
Email ID: [parthsutariya31@gmail.com](mailto:parthsutariya31@gmail.com)  
Country: India  
College: BSBI, Berlin, Germany  
Specialization: Data Analyst

### Problem Description:

---

- In the fourth week of the project, we delved into exploratory data analysis (EDA) to gain insights into the characteristics and distribution of the data.
- The objective is to understand the underlying patterns, trends, and challenges in the dataset, which are crucial for developing an effective hate speech detection model.
- EDA helps us identify potential biases, class imbalances, and anomalies in the data, guiding preprocessing and modeling decisions to improve the performance and fairness of the model.

### EDA Process:

---

- **Data Overview:** Start by obtaining a high-level overview of the dataset, including its size, structure, and basic statistics.
- **Class Distribution:** Analyze the distribution of hate speech and non-hate speech instances to assess class balance or imbalance.
- **Text Length Distribution:** Investigate the distribution of text lengths (e.g., character count, word count) in hate speech and non-hate speech categories.
- **Word Frequency Analysis:** Identify the most frequent words or phrases in hate speech and non-hate speech categories to understand common patterns and language usage.
- **Sentiment Analysis:** Conduct sentiment analysis to explore the overall sentiment of text in hate speech and non-hate speech instances.